EXHIBIT A

Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development

John J. Mekalanos, Daryl J. Swartz & Gregory D. N. Pearson

Department of Microbiology and Molecular Genetics, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

Nigel Harford, Francoise Groyne & Michel de Wilde

Department of Molecular Genetics, Smith-Kline-R.I.T., rue de l'Institut, 89, B-1330 Rixensart, Belgium

Nucleotide sequence and deletion analysis have been used to identify the regulatory and coding sequences comprising the cholera toxin operon (ctx). *Incorporation of defined* in vitro-*generated* ctx *deletion mutations into* Vibrio cholerae *by* in vivo *genetic recombination produced strains which have practical value in cholera vaccine development.*

MODERN history has recorded seven world pandemics of cholera, a diarrhoeal disease produced by the Gram-negative bacterium *Vibrio cholerae*[1]. Laboratory tests can distinguish two biotypes of *V. cholerae*, classical and El Tor, the latter being responsible for the most recent cholera pandemic. The diarrhoeal syndrome induced by colonization of the human small bowel by either biotype of *V. cholerae* is caused by the action of cholera toxin, a heat-labile enterotoxin secreted by the growing vibrios[2]. Diarrhoeal diseases affecting both humans and animals caused by some enterotoxinogenic strains of *Escherichia coli* are also induced by a heat-labile enterotoxin (LT) which is closely related to cholera toxin in structure and mode of action[1-3].

Cholera toxin is an 84,000-molecular weight (MW) protein composed of one A subunit (27,000 MW) and five B subunits (11,600 MW). The A subunit, although synthesized as a single polypeptide chain, is usually proteolytically nicked to form two disulphide-linked polypeptides, A1 (22,000 MW) and A2 (5,000 MW)[4,5]. The A1 polypeptide is an enzyme and promotes the activation of adenylate cyclase in target cells by catalysing the ADP-ribosylation of a GTP-binding regulatory component of the cyclase complex[6]. The resulting accumulation of cyclic AMP in the intestinal mucosa leads to the severe fluid loss characteristic of cholera. Each B subunit has a high binding affinity for the toxin's cell surface receptor, ganglioside GM_1 (ref. 7). Neutralizing antibodies raised against the holotoxin react mainly with the B subunits[1,2].

Much of the current interest in the genetics of cholera toxin has been promoted by the need to develop a more efficacious vaccine against this enterotoxic disease. Parenterally administered, killed whole-cell and toxoid vaccines have been shown to be largely ineffective in producing long-lasting immunity to cholera, presumably because they lack the ability to induce local immune responses in the intestine[8,9]. Since the natural disease is capable of inducing prolonged immunity[9,10], several investigators have proposed the use of attenuated, non-toxinogenic mutants of *V. cholerae* as live oral cholera vaccines[11-16]. While encouraging results in volunteer studies have been obtained with some of these strains, factors such as genetic instability or poor colonizing ability have contraindicated their use in the field[15-17].

The recent relaxation of US governmental guidelines prohibiting the molecular cloning of bacterial toxin genes has permitted the use of a powerful new approach to the analysis of cholera toxin gene structure and vaccine development. These studies have shown that like the *elt* genes, which encode *E. coli* LT[18], the genes for the A and B subunits of cholera toxin are arranged in a single transcriptional unit with the A cistron (*ctxA*) preceding the B cistron (*ctxB*)[19]. *V. cholerae* strains of the classical biotype contain a nontandem, chromosomal duplication of the *ctx* operon that is structurally identical in all strains.[53] In contrast, about 70% of El Tor strains have only a single copy of *ctx*, while the remaining strains have two or more *ctx* copies present on a tandemly repeated genetic element. This genetic duplication and amplification of the toxin operon may be related to the instability observed in some of the earlier *V. cholerae* toxin mutants[13,16].

In this article, we report the entire nucleotide sequence of one *ctx* operon together with partial sequences containing the *ctx* promoter regions of five other cloned *ctx* copies. Deletion analysis has allowed the identification of toxin transcriptional and translational regulatory sequences. An *in vitro*-constructed, internal deletion in *ctxA* was recombined *in vivo* into both *ctxA* gene copies of *V. cholerae* strain Ogawa 395. Since this genetic recombinant still produces the immunogenic B subunit of the toxin, it should have practical value in cholera vaccine development.

Molecular cloning of *ctx* operon copies

A total of six *ctx* copies were cloned from four *V. cholerae* strains. These include both *ctx* copies from strain 569B, both copies from strain RV79, one of two copies from strain E7946 and the single *ctx* copy of strain 2125. With the exception of the classical strain 569B, all these strains are El Tor in biotype. The *ctx* copies were cloned as various *V. cholerae* restriction fragments which hybridized with ^{32}P-labelled *elt* or *ctx* probes as described in the legend to Fig. 1.

The restriction sites for several endonucleases were located on these cloned inserts, and the resulting maps were aligned at the conserved *Xba*I site previously determined to lie early in the A cistron[19], (Fig. 1*A*). Other conserved restriction sites for *Nru*I, *Pst*I, *Ava*I and *Bgl*II were also found preceding *ctxA* on these various inserts (Fig. 1). Additional restriction mapping and hybridization analysis has indicated that the 5 kilobase pairs (kbp) of DNA directly preceding the toxin structural genes is the same for all cloned *ctx* copies thus far examined[53]. Since the larger chromosomal sequence environment flanking these different *ctx* copies appears to vary as determined by Southern blot hybridization, we have proposed that the conserved DNA immediately upstream of *ctx* is part of a genetic element responsible for toxin operon duplication and transposition events. From a practical point of view, the conserved 5' flanking sequences associated with *ctx* provided part of the necessary homology for efficient *in vivo* recombination of *in vitro* constructed deletion mutations into multiple copies of the toxin operon (see below).

Nucleotide sequence of *ctx*

The strategy used to determine the complete nucleotide sequence of the single *ctx* copy of strain 2,125 is shown below the restriction map of the insert cloned on pRIT10824 (Fig. 1*B*). The 2,020 nucleotides determined are shown in Fig. 2. Comparison of the 2,125 nucleotide sequence with both the *elt* nucleotide sequence[20,21] and known amino acid sequences of

0028-0836/83/490551—07$01.00

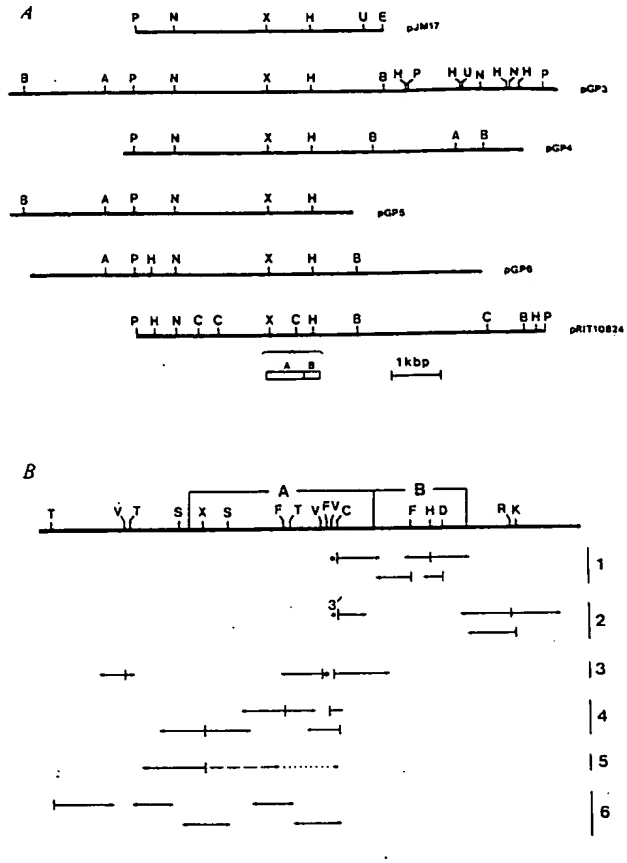

Fig. 1 *A*, Restriction maps of cloned *V. cholerae* DNA fragments carrying *ctx* genes. The maps are aligned over the common *Xba*I restriction site in the A subunit coding sequence. The limits of the coding regions for the A and B subunits are shown below the maps while the names of plasmids carrying the various inserts are shown to the right. pJM17 and pGP3 each contain, respectively, the two different *ctx* copies from strain 569B. pGP4 and pGP5 contain the two different *ctx* copies from strain RV79, and pGP6 contains one of two *ctx* copies from strain E7946. pRIT10824 contains the single *ctx* copy cloned from strain 2125. *B*, Enlarged map and sequencing strategy used to determine the sequence of the 2125 *ctx* operon. The upper line shows an expanded restriction map covering the junction between the 1.6 kbp and 3.9 kbp *Cla*I fragments on pRIT10824. The location of the A and B subunit coding regions is indicated above the line, while the arrows below denote the sites, direction and extent of sequencing. The following letters are used to denote different restriction enzyme sites on the maps; A, *Ava*I, B, *Bgl*II; C, *Cla*I; D, *Dde*I; E, *Eco*RI; F, *Hin*fI; H, *Hin*cII; K, *Hae*III; N, *Nru*I; R, *Rsa*I; S, *Sau*3A; T, *Taq*I; U, *Pvu*I; V, *Hpa*II; X, *Xba*I. **Methods:** Cloning methods. The construction of plasmid pJM17[19] involved the cloning of an *elt*-homologous *V. cholerae* *Pst*I–*Eco*RI restriction fragment recovered by electroelution from agarose gel slices. A similar strategy was used to clone the single *ctx* copy from strain 2125 into pBR322[41] as three independent fragments: a 1.6 kbp *Cla*I fragment (in pRIT10841) hybridizing to an *eltA* gene probe, a 3.9 kbp *Cla*I fragment (in pRIT10810) hybridizing to an *eltB* gene probe, and an 8.3 kbp *Pst*I fragment (in pRIT10824) hybridizing to the 1.6 kbp *Cla*I fragment of pRIT10841. The *eltA* and *eltB* gene probes and hybridization conditions were those previously described[42]. Only the map for the 8.3 kbp insert of pRIT10824 is shown here. The other *ctx* inserts cloned on plasmids pGP3, pGP4, pGP5 and pGP6 were isolated from genomic libraries of strains 569B, RV79 and E7946 prepared by insertion of *Sau*3A-partially digested chromosomal fragments into the *Bam*HI site of plasmid pBR327[41] (see below). Only partial restriction maps of the inserts present on pGP3 and pGP5 are shown in this figure. **Construction of genomic libraries:** About 250 μg of *V. cholerae* DNA was partially digested with *Sau*3A (New England Biolabs) by incubation in the presence of 33 units of enzyme at 37 °C for 10 min in a buffer containing 6 mM Tris-HCl, *p*H 7.5, 5 mM $MgCl_2$ and 50 mM NaCl, followed by heat inactivation at 70 °C for 10 min. The digested DNA was fractionated by velocity sedimentation through a 10–40% sucrose gradient, and fractions containing 7–12 kbp fragments were precipitated by addition of two volumes ethanol. Approximately 1 μg of this partially digested DNA was mixed with 2 μg of pBR327[41] which had been previously digested with *Bam*HI (New England Biolabs) and bacterial alkaline phosphatase (Bethesda Research Laboratories). The DNA mixture was incubated at 16 °C for 6 h, in 300 μl of a solution containing 30 mM Tris-HCl, *p*H 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 0.4 mM ATP, 5 mM dithiothreitol, and 600 units of T_4 DNA ligase. Transformation of this ligation mixture into *E. coli* MS371 (K-12 F^- *gal thi endA sbcB hsdR4 hsdM*$^+$) yielded a library consisting of $0.5–2 \times 10^4$ ampicillin resistant transformants, of which 90% were tetracycline sensitive and carried recombinant plasmids with 7–11 kbp *V. cholerae* DNA inserts. Libraries were screened by colony hybridization using ^{32}P-labelled *eltA* and *eltB* gene probes and hybridization conditions as described previously[19,42]. **DNA sequencing:** Maxam and Gilbert sequencing[43] was done on the following plasmids: (1) pRIT10810, (2) pRIT10809, same as pRIT10810 but with its 3.9 kbp *Cla*I insert in the reverse orientation, (3) pRIT10824 or (4) pRIT10841. The sequencing was performed on 5′-labelled fragments or in the one case indicated a 3′-labelled fragment from pJM17, pGP3, pGP4, pGP5 and pGP6 as follows: the dotted line denotes the extent of sequencing done on 3′-labelled fragments derived from pJM17, the dashed line for fragments from pJM17 and pGP3, and the solid line for fragments from plasmids pJM17, pGP3, pGP4, pGP5 and pGP6. An asterisk denotes that labelling was at an adjacent site on the plasmid vector. Polynucleotide kinase and the Klenow fragment of DNA polymerase I were used, respectively to label the 5′ and 3′ ends of fragments essentially as described elsewhere[36]. Sequencing using M13[6] was done by the dideoxy method[44] after subcloning *Taq*I or *Sau*3A fragments from pRIT10824 or pRIT10841 into the M13mp7 vector[37]. A synthetic, phage-specific primer was used to initiate complementary strand synthesis.

569B-derived cholera toxin subunits[22–26], shows that nucleotides 516 to 1,292 and 1,289 to 1,663 form the coding sequences for *ctxA* and *ctxB*, respectively. The primary translation product of *ctxA* is a 258 amino acid long polypeptide, while that of *ctxB* is a 124 amino acid polypeptide. As in the case of LT[20,21], both polypeptides are apparently precursors with 18 and 21 amino acid hydrophobic, amino-terminal signal sequences. The A2 coding sequence lies at the carboxy-terminal end of the A subunit. The calculated molecular weights of the mature subunits are 21,817 for A1, 5,398 for A2, and 11,677 for B.

Comparison of the DNA-deduced 2,125 protein sequences, with polypeptide sequences[22–26] determined for the 569B toxin subunits, revealed some major differences. Since partial sequencing of the two 569B *ctxA* copies showed no changes with respect to the 2,125 sequence from nucleotide 413 to 830 for the pGP3 *ctxA* copy and from 413 to 1,194 for the pJM17 copy, the discrepancies found in this region are most likely mistakes in amino acid sequencing. The agreement with the A2 polypeptide sequence is good except for an Ile–Asp inversion (residues 229–230) and an Asp–Asn transition (residue 238). The five differences between the deduced polypeptide sequence of the B subunit of 2,125 toxin and the sequence determined for the corresponding 569B polypeptide[25,26] could be the result of single base changes.

Comparison of *elt* and *ctx* revealed that at the nucleotide level, A and B cistrons of the two different toxin operons are 75% and 77% homologous. This percentage is reflected at the amino acid level (76% and 78% homology for A and B polypeptides, respectively). The amino acid sequences have been largely conserved, with the differences scattered throughout the sequence except for the region around the cleavage site between A1 and A2 where the homology drops to 33% between amino acids 189 and 212. Also, the A subunit of LT is four amino acids shorter than the cholera toxin A subunit.

With respect to noncoding regions, only 90 nucleotides upstream of the *eltA* initiation codon are available for comparison. In marked contrast to the coding sequences, the *elt* and *ctx* operons have diverged completely in these potential promoter regions. With regard to transcriptional termination, nucleotides 1,669–1,694 of the *ctx* sequence exhibit dyad symmetry (Fig. 2) with a G+C-rich stem–loop and a poly(T) tail depending on the length of the stem. These features are characteristic of prokaryotic, *ρ*-independent transcription termination signals[27,28] and suggest that *ctxA* and *ctxB* are the only genes in the *ctx* operon.

Translational signals in *ctx* expression

The *ctxA* gene has two overlapping Shine–Dalgarno (SD) sequences[29] of which GGAG is the more correctly spaced, relative to the ATG start of the structural gene (Fig. 3). The *ctxB* gene exhibits a theoretically perfect SD sequence (TAAGGA) which lies within the carboxy terminal coding

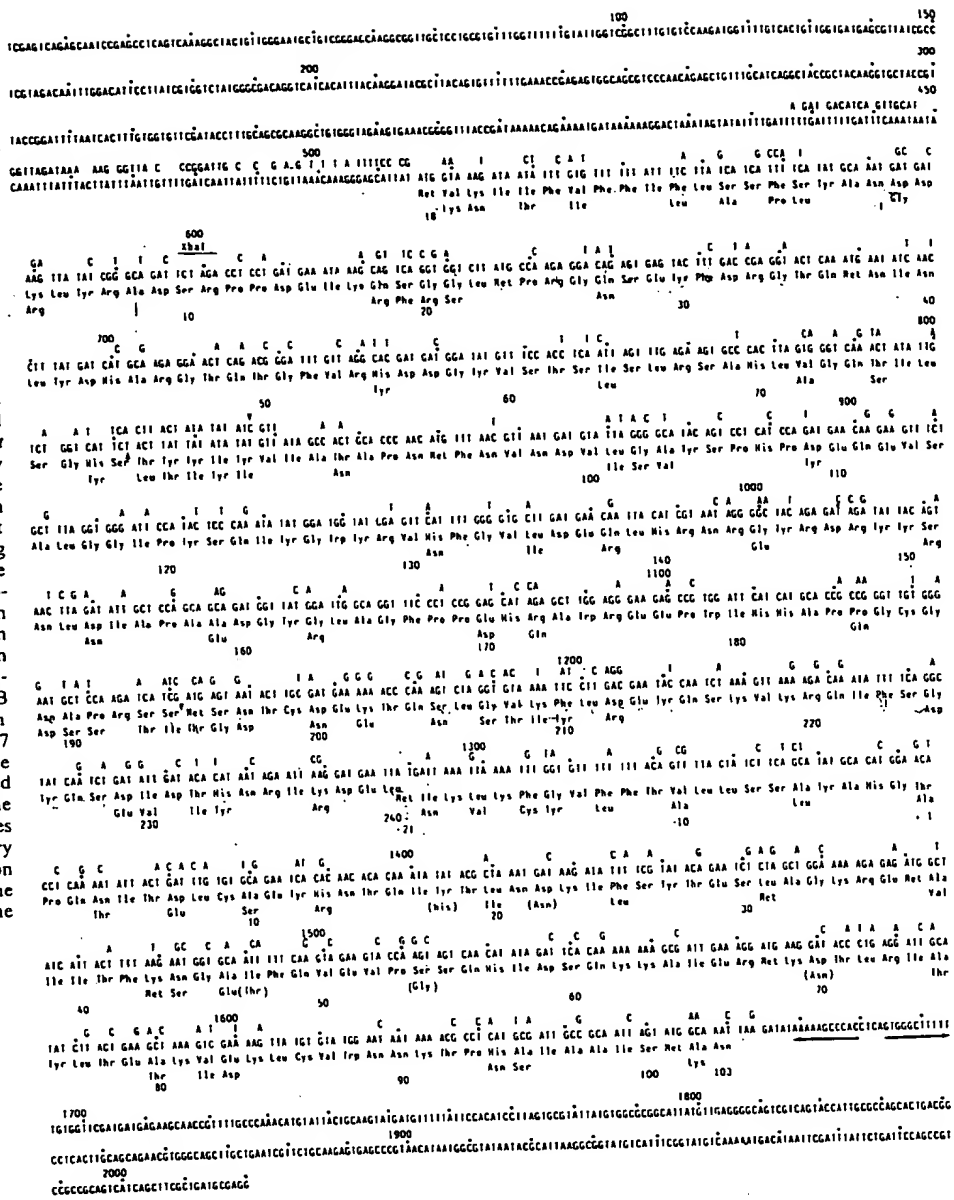

Fig. 2 DNA sequence of the *V. cholerae* toxin operon from strain 2125. The antisense strand is shown from 5' to 3'. From nucleotide 427 to 1,663, the sequence is compared with the published sequence of LT genes[20,21]; *elt* nucleotides are shown above the sequence only where they differ from the *ctx* sequence except between nucleotides 810 and 830, where deletion of a T (arrowed) in the *elt* sequence creates a frameshift which is corrected by insertion of a C (arrowed) 16 bp downstream (see text). Analogous events have previously been seen after pseudoreversion of frameshift mutations[45] but to our knowledge, this is the first naturally occurring case described. The deduced amino acid sequence of *ctxA* (nucleotide 516 to 1,289) and *ctxB* (nucleotide 1,289 to 1,660) is shown and compared with that of LT. Amino acids that differ in LT are shown below the cholera toxin amino acid sequence. In addition, for the mature B subunit sequence, differences from the published amino acid sequence of B subunit purified from strain 569B[26,27] are shown in brackets. Two of these differences (amino acids 47 and 54) are also found in LT. The cleavage site between the A1 and A2 polypeptides is indicated by an arrow (amino acids 194–195). Note also the overlap of *ctxA* and *ctxB* cistrons (nucleotides 1,289–1,292). Sequence exhibiting dyad symmetry and potentially involved in transcription termination is indicated with divergent arrows. Features of the sequence immediately upstream of the *ctxA* gene are detailed in Fig. 3.

sequence (nucleotides 1,277–1,282, Fig. 2) of the *ctxA* cistron. The first two nucleotides of the *ctxA* translation termination signal TGA are the last two nucleotides of the *ctxB* translation initiation triplet ATG. This particular overlapping arrangement is also found several times in phage λ operons[30] and may be involved in translational coupling[31] of the *ctxA* and *ctxB* genes. However, evidence presented below suggests that this is not the case with the *ctx* operon. Where documented, translational coupling is observed between cistrons whose gene products interact in a one to one stoichiometry[31], and in contrast, the cholera toxin molecule is composed of one A subunit and five B subunits. Moreover, *E. coli* produces stoichiometrically 7 times more cholera toxin B subunit than A subunit (data not shown). Fusion of the *ctxB* gene to various *E. coli* promoters allows high expression of *ctxB* in the absence of *ctxA* translational initiation signals. These data suggest that translation of *ctxB* relies primarily on independent initiations promoted by its own ribosome binding site.

Another experiment supports this conclusion. Our DNA sequencing analysis identified two *Nde*I sites at positions 561 and 1,337 within the *ctxA* and *ctxB* genes, respectively. The positions of these sites relative to the reading frames of *ctxA* and *ctxB* allowed us to construct a *ctxA* deletion which codes for an in-frame fusion of amino acid 17 of the A subunit signal sequence to amino acid 19 of the B signal and thus maintains the normal processing site of the B signal sequence (residue 21). This genetic fusion makes B subunit expression dependent on the efficiency of the A cistron translation initiation sequences, provided the hybrid signal sequence is processed at normal efficiency. *Nde*I digestion of plasmid pGP3 followed by ligation produced such a fusion between these two sites and gave plasmid pJM3.1. Plasmid pJM3.1 produced 0.056 μg ml^{-1} of B subunit in *E. coli* MS371 while pGP3 produced 0.50 μg ml^{-1}. These data suggest that the *ctxB* ribosome binding site is about ninefold more efficient than the *ctxA* site.

Toxin promoter regions

We determined approximately 200 base pairs of sequence upstream of the *Xba*I sites for each of the other five additional cloned copies of the *ctxA* gene, cloned on plasmids pGP3, pGP4, pGP5, pGP6 and JM17. Comparison of these sequences with the corresponding region of the *ctxA* gene derived from strain 2,125 indicated a perfect conservation of sequence between these copies from nucleotides 413 to 590 with one notable exception. The sequence TTTTGAT comprising nucleotides 419–425, 426–432 and 433–439 of the 2,125 sequence was found tandemly repeated 3–8 times preceding different *ctxA* gene copies (Fig. 3). Figure 3 shows part of a sequencing gel autoradiograph that spans DNA carrying eight of these tandem repeats in the region adjacent to the *ctxA* gene of pJM17.

To determine the position of the toxin operon promoter with respect to these repeated sequences, we used nuclease *Bal*31

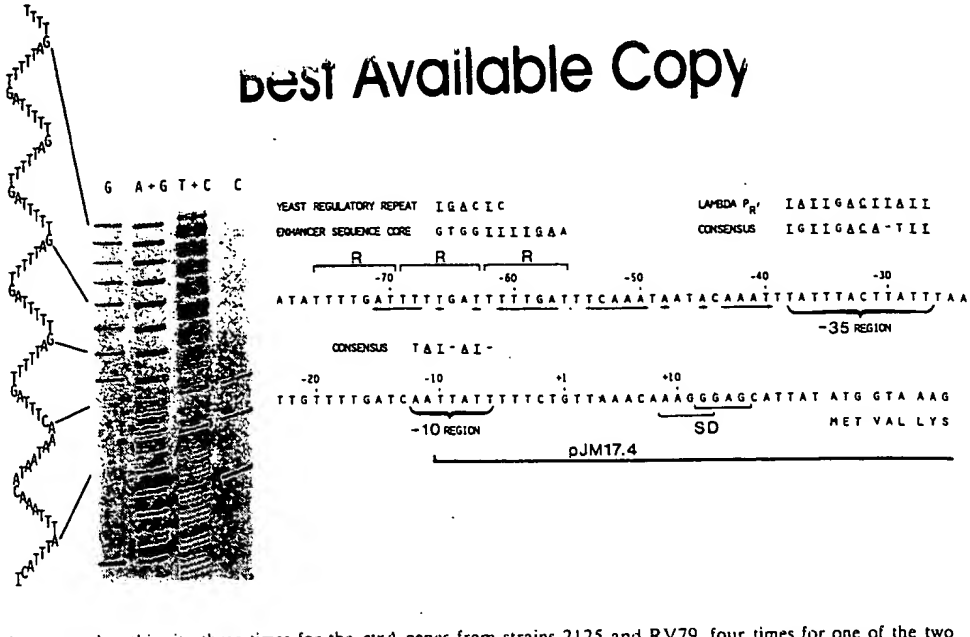

Fig. 3 DNA sequence preceding the start of the *ctxA* cistron. The sequence of the antisense strand of DNA corresponding to nucleotides 416 to 524 of the 2125 sequence is shown. The sequence is numbered relative to the proposed +1 mRNA start point at nucleotide 495. Sequences constituting the −10 and −35 regions of the proposed toxin promoter are indicated. Above these are consensus sequences determined by Pribnow[46] and Rosenberg and Court[27] to be present, with some variation, in corresponding regions of most *E. coli* and virus promoters. The −35 region of the λ P'_R promoter[32] is also shown for comparison. Nucleotides underlined in the consensus and P'_R sequences are also present in the proposed *ctx* promoter sequence. SD refers to nucleotides which are complementary to sequences located near the 3′ terminus of 16S rRNA[29]. The DNA removed (+ or − two nucleotides) from this region by the deletion carried by plasmid pJM17.4 is indicated by the lower heavy line. Divergent arrows under the indicated nucleotides designate regions of hyphenated dyad symmetry. Three copies of the tandemly repeated sequence TTTTGAT are indicated (R). This sequence was found tandemly repeated at this site three times for the *ctxA* genes from strains 2125 and RV79, four times for one of the two *ctxA* copies from strain E7946 and eight times for each of the *ctxA* copies from strain 569B. The figure also shows that this repetitive region exhibits nucleotide sequence homology with a repetitive element present in Ig and viral enhancer sequences[47] and the core sequence of a yeast regulatory repeat[48]. To the left of the nucleotide sequence is part of a sequencing gel autoradiograph showing the region 5′-proximal to the *ctxA* gene of pJM17 which carries eight copies of the TTTTGAT repeat.

to construct a set of deletions around the *Xba*I site on plasmid pJM17. Expression of the *ctxB* gene was used to score the effects of these deletions on toxin operon transcription in *E. coli* MS371 and the *V. cholerae* toxin deletion mutant M7922[15]. The smallest deletion upstream of the *ctxA* gene which substantially reduced B subunit production in both of these bacterial backgrounds was carried by pJM17.4. The inhibitory effect of the deletion carried by pJM17.4 was shown to be due to the small 130 bp deletion 5′-proximal to the *Xba*I site of this plasmid by analysing the expression of the reassorted plasmids pJM23, pJM31 and pJM32 (Table 1).

Figure 3 shows the sequence of the antisense strand of DNA 5′-proximal to the *ctxA* gene and the extent of DNA removed in this region by the deletion on pJM17.4. The deletion begins within *ctxA* and extends about 31 bp upstream from the start of the A cistron. Since this deletion greatly reduced expression of the B gene, it appears that some essential component of the *ctx* promoter is located within the 31 nucleotides immediately preceding the initiation codon of *ctxA*. This region contains a sequence AATTATT, which matches four of the five specific bases of the Pribnow consensus sequence, TATNATN. Given that most transcripts initiate at the first purine residue 5–7 nucleotides downstream from the Pribnow box[27], we have tentatively assigned the G residue located at position 495 in the 2,125 sequence as the +1 position of the toxin transcript (Fig. 3).

Our proposed −10 region is associated with a −35 region that shared 7 of 11 bases with the consensus sequence for this region[27] and is identical to the −35 region of the strong P'_R promoter of phage λ[32] except for one G to T base change. Directly preceding the proposed −35 region of the toxin promoter lies a 32 bp region of hyphenated dyad symmetry (Fig. 3). The position of this symmetrical sequence relative to the −35 region suggests that it may have a role in positive regulation[27] of the toxin promoter. In regard to this possibility, we have recently cloned a *V. cholerae* regulatory gene which activates *ctx* transcription approximately 100-fold in *E. coli* (V. Miller and J.M., manuscript in preparation). Another characteristic of the symmetrical region is that half of its structure is contributed by DNA containing the tandemly repeated sequence TTTTGAT. It is of interest that strain 569B, which carries eight of these repeats upstream of each of its two toxin operon copies, produces about two orders of magnitude more toxin than strains 2,125, RV79 and E7946. However, the different toxin operon copies cloned from these four strains all express about the same amount of B subunit in *E. coli*. Therefore, if the repetitive region is important in determining differences in toxin expression between *V. cholerae* strains, then these sequences probably interact with regulatory components unique to *V. cholerae*. In this respect, a 30–50-fold increase in B subunit production for plasmids pJM20, pJM23 and pJM32 was observed in *V. cholerae* M7922 as compared with *E. coli* MS371 (Table 1). Since these plasmids all express B via the natural toxin promoter, and actually have copy numbers 3–4-fold lower in M7922 than in MS371 (as determined by hybridization), these data suggest that expression of the toxin operon is about 100-fold higher in *V. cholerae* than in *E. coli*.

Construction of *ctx* deletion derivatives

Ogawa 395 (0395) is a classical strain of *V. cholerae* which has been shown to induce in human volunteers protective immunity against homologous or heterologous serotypic rechallenge that was greater than 3 yr in duration[9,10,17]. We have constructed isogenic A^-B^+ and A^-B^- toxin derivatives of 0395 by *in vivo* recombination of *in vitro* constructed deletions of the *ctx* operon into each of the two resident *ctx* copies carried by this strain.

Plasmid pJM23 served as the starting point in these strain constructions since it carries an internal *ctxA* deletion which by DNA sequencing was shown to span nucleotides corresponding to 597 to 1,059 of the 2,125 sequence (Fig. 2). This deletion removes over 80% of the DNA sequence required for formation of the A1 polypeptide and all of the sequence corresponding to residues thought to be involved in formation of the A1 active site[33]. A second *ctx* deletion was constructed which removes the DNA between the *Xba*I site at nucleotide 597 and the *Hin*cII site at nucleotide 1,515, thus deleting essentially all of *ctxA* and half of *ctxB* (Fig. 2). DNA fragments carrying the *ctx* deletions were subcloned as *Eco*RI fragments into plasmid pRK290[34] to give plasmids pJM290.2 and pJM290.211, respectively (Fig. 4*A*): The 918 bp *ctxB* deletion present on pJM290.211 is actually a deletion-substitution since a 1.4 kbp DNA fragment encoding resistance to Kanamycin (Km) was inserted at the position formally occupied by the deleted *ctx* DNA (Fig. 4*A*).

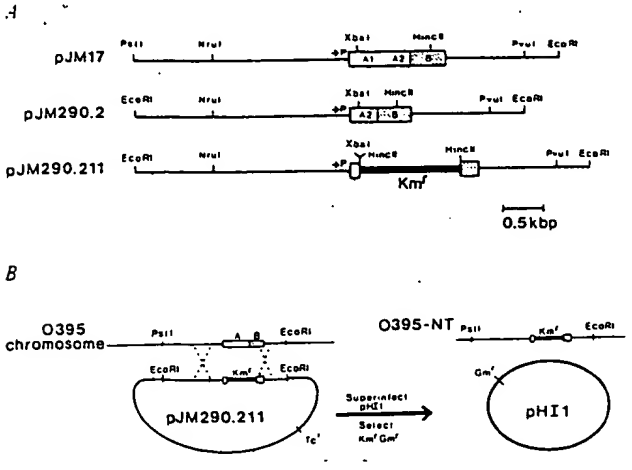

Fig. 4 *A*, Restriction maps of *ctx* inserts carried by pJM17, pJM290.2 and pJM290.211. *B*, Marker exchange procedure used to construct strain 0395-NT.

Methods: Plasmid construction. Exonuclease *Bal*31 treatment of pJM17 followed by ligation in the presence of synthetic *Xba*I DNA linkers was used to construct plasmid pJM17.23 as described in Table 1. The reassorted plasmid pJM23 was constructed by replacing the 2.1 kbp *Pst*I–*Xba*I fragment of pJM17.23 with the original 2.7 kbp *Pst*I–*Xba*I fragment of pJM17 which carries the *ctx* promoter (arrow marked P). The *Pst*I site of pJM23 was converted to blunt ends by treating 0.3 μg of *Pst*I digested pJM23 with 1 unit of T4 DNA polymerase (Bethesda Research Labs) for 10 min at 37 °C in a buffer containing 33 mM Tris-acetate, *p*H 8.0, 66 mM Na acetate, 10 mM Mg acetate, and 0.2 mM of each of the four deoxynucleotide triphosphates. Subsequent ligation in the presence of a 50-fold molar excess of synthetic *Eco*RI linker (Collaborative Research) followed by transformation and selection of tetracycline resistance gave plasmid pJM23.2. The 4.6 kbp *Eco*RI insert of pJM23.2 was then subcloned into the *Eco*RI site of pRK290[34] to give pJM290.2. A 1.4 kbp *Eco*RI fragment from pUC71K (gift from J. Vieira and J. Messing) encoding resistance to kanamycin (Km) and also flanked by two *Hin*cII sites, was modified with *Xba*I linkers and inserted in the *Xba*I site of pJM23.2 to give pJM23.21. Digestion of pJM23.21 with *Hin*cII, followed by ligation and selection of Km^r Tc^r transformants gave pJM23.211. The 5.6 kbp *Eco*RI insert of pJM23.211 was then subcloned into the *Eco*RI site of RK290 to give pJM290.211. The marker-exchange procedure of Ruvkun and Ausubel[34] was used to recombine the *in vitro*-constructed *ctx* deletion mutation carried by pJM290.211 onto the chromosome of *V. cholerae* 0395 *in vivo*. Mobilization of pJM290.211 into a spontaneous Sm^r derivative of 0395 was performed essentially as described in the legend to Table 1 for strain M7922 and pJM17 derivatives. Transconjugants which were Sm^r Tc^r Km^r carried exclusively pJM290.211 as evidenced by the nontransmissible nature of the Km^r and Tc^r markers. Superinfection with the gentamycin (Gm) resistance plasmid pHI1 (same as pPH1JI)[49,50] was performed in similar plate matings between CSH56/pHI1 and 0395/pJM290.211 by selecting transconjugants on LB agar containing 30 μg ml^{-1} Gm, 30 μg ml^{-1} Km and 100 μg ml^{-1} Sm. Since pHI1 and pJM290.211 are both P-group plasmids and are therefore incompatible, the Gm^r Km^r Sm^r Tc^s transconjugants obtained had recombined the Km^r of pJM290.211 onto the 0395 chromosome via crossover events (dashed lines) occurring between flanking homologous regions of one of the resident chromosomal *ctx*$^+$ copies and the corresponding DNA on the insert of pJM290.211. Thus, the *ctx*-Km^r deletion-substitution mutation carried by pJM290.211 replaced the normal *ctx* locus as shown (*B*). However, since 0395 has two copies of the toxin locus, it was necessary to recombine the *ctx*-Km^r deletion-substitution mutation into the other copy as well. This was accomplished by growing cells containing one marker-exchanged copy for several generations during which time homologous sequences flanking the other *ctx* copy allowed recombination events to transfer the *ctx*-Km^r deletion-substitution mutation into the other *ctx* copy. 0395 derivatives carrying the *ctx*-Km^r deletion-substitition mutation in both *ctx* copies occurred at a frequency of 10^{-2} and were recognized by colony hybridization[51] using a probe (CT-1, Fig. 5) which specifically hybridizes to the DNA removed by the deletion mutation. One of these double marker-exchanged derivatives was subcultured overnight in the absence of Gm and underwent spontaneous curing of pHI1 to give strain 0395-NT. The construction present on pJM290.2 was crossed onto the chromosome by a modified marker-exchange procedure. Plasmid pJM290.2 was mobilized into strain 0395-NT which had both of its two *ctx* copies tagged with Km^r. Introduction of pJM290.2 into this strain allowed cross-over events, analogous to those shown in section B, to subsequently replace both Km^r-tagged copies with the simple *ctxA* deletion present on pJM290.2. After overnight subculturing of 0395-NT/pJM290.2 in the absence of Km, these double recombinant colonies were recognized as spontaneous Km^s segregants, which arose at a frequency of 10^{-3} when single colonies were analysed by replica plating. One of these was spontaneously cured of pJM290.2 by subculturing in the absence of Tc to give strain 0395-N1. Southern blot hybridizations (Fig. 5) were used to confirm the expected structures of the 0395-NT and 0395-N1 *ctx* loci.

The marker exchange procedure of Ruvkun and Ausubel[34] was used to recombine the deletion-Km^r substitution mutation carried by pJM290.211 into one of the *V. cholerae* 0395 chromosomal *ctx* copies (Fig. 4*B*). Subsequent *in vivo* recombination spontaneously produced a genetic recombinant (0395-NT) that had both of its resident *ctx* copies replaced with the *ctx* deletion-Km^r substitution mutation carried by pJM290.211. A second *ctx* marker exchange step with pJM290.2 replaced both Km^r tagged *ctx* copies of 0395-NT with the internal *ctxA* deletion construction carried by pJM290.2 to give strain 0395-N1.

Southern blot hybridization, using *Xba*I genomic digests and probes derived from the *elt* and *ctx* genes, was used to confirm the genetic structures of 0395-N1 and 0395-NT. The two *ctx* loci of the parental strain 0395 were seen as two bands (9.8 and 7.9 kbp) which reacted with all three probes used in the analysis (Fig. 5). In contrast, 0395-N1 and 0395-NT were both unreactive with a probe, LT-A1, specific for the A1 region of *ctxA*. The CT-1 probe, which exactly corresponds to the *ctx* DNA deleted on pJM23.211, did not react with 0395-NT but did hybridize to two restriction fragments of 0395-N1. Each of these fragments was smaller than the corresponding 0395 band by the size of the *ctxA* deletion originally present in pJM23 (~450 bp). The CT-B1 probe, derived from DNA downstream from the *Hin*cII site in *ctxB*, hybridized to the same bands as the CT-1 probe for 0395-N1. The CT-B1 probe reacted with two bands in 0395-NT which were each 500 bp larger than the corresponding two bands of 0395. This is the difference in size between the 0.9 kbp *ctx* deletion carried on pJM290.211 and the 1.4 kbp Km^r fragment inserted at the deletion junction.

These results, together with analogous results obtained with restriction enzymes other than *Xba*I, confirmed that 0395-N1 and 0395-NT carry the appropriate *ctx* deletions mutations in both of their two *ctx* copies. Consistent with this conclusion, 0395-N1 produces the same amount of the B subunit (0.3 μg ml^{-1}) as the 0395 parental strain but displays no toxicity detectable in Chinese hamster ovary cells[35], while 0395-NT produces neither the A subunit nor the B subunit of cholera toxin (data not shown).

Discussion

Our determination of the cholera toxin nucleotide sequence has confirmed the close evolutionary relationship of *ctx* and *elt*. However, the nucleotide sequence homology observed is confined entirely to the A and B polypeptide coding sequences. The 5′ control regions upstream from these structural genes have completely diverged for the two toxin operons. We also found that the *ctx* operon expressed much more efficiently in *V. cholerae* than in *E. coli*. Indeed, genetic fusions of the very strong *tac* promoter[36] to *ctxB* actually produce only one-third as much B subunit in *V. cholerae* as the natural *ctx* promoter derived from strain 569B (J.J.M., unpublished results). These observations suggest that the structure of the *ctx* promoter region reflects its interaction with regulatory proteins unique to *V. cholerae*. The existence of a variety of *V. cholerae* toxin regulatory mutations supports this hypothesis[37,38]. In contrast, the *elt* operon, being located extrachromosomally on plasmids[2], may have been forced to evolve more universal transcriptional signals which would function reasonably well in a variety of Gram-negative backgrounds.

Although we have not yet identified the function of the tandemly repeated sequence TTTTGAT found in the *ctx* promoter region, we point out that such short repetitive structures have not been previously observed in a prokaryotic organism. It is also of interest that much larger tandem duplications of regions either adjacent to or encompassing the entire *ctx* operon are also very common among strains of *V. cholerae*[53]. Thus, *V. cholerae* may have evolved several different levels of genetic control which involve DNA amplification.

In addition to allowing the determination of the *ctx* nucleotide sequence, molecular cloning has also allowed us to construct

Table 1 Effects of *Bal*31 deletions on B subunit expression

Plasmid	Size of deletion upstream of *Xba*I site	Size of deletion downstream of *Xba*I site	B subunit production *E. coli* MS371	B subunit production *V. cholerae* M7922
pJM17	0	0	0.3	ND
pJM17.4	0.13	0.40	<0.005	0.2
pJM32	0	0.40	0.3	10.0
pJM31	0.13	0.45	<0.005	0.2
pJM23	0	0.45	0.2	9.0
pJM20	0	0.30	0.2	10.0

Plasmid construction: five μg of pJM17 DNA[19] was linearized with *Xba*I and treated with 6 units of exonuclease *Bal*31 (New England Biolabs) for 2–60 min. The cut-back plasmid was phenol extracted, ethanol precipitated, and then ligated in the presence of a 50-fold molar excess of synthetic *Xba*I linker (New England Biolabs). The plasmid was redigested with *Xba*I, ligated and transformed into *E. coli* MS371 selecting for tetracycline resistance. Deletion plasmids were characterized by digesting with *Pst*I and *Xba*I to excise the fragment upstream of the original *Xba*I site of pJM17 or with *Xba*I and *Eco*RI to excise the fragment downstream of the original *Xba*I site of pJM17 (see restriction map, Fig. 1). The size of deletions in these fragments when compared to the original fragments is shown in kilobase pairs. Reassorted plasmids (pJM20, pJM23, pJM31 and pJM32) were constructed by replacing the *Pst*I–*Xba*I or *Xba*I–*Eco*RI fragment on one deletion plasmid with the analogous fragment from other deletion plasmids (pJM17.20 or pJM17.23, not shown) or from pJM17, as indicated by the size of the deletions. Plasmids were mobilized into the *V. cholerae* toxin deletion mutant M7922[15] using the plasmid pRK2013[34] as follows: *E. coli* MS371 carrying various pJM17 derivatives was grown in LB broth containing 15 μg ml^{-1} tetracycline (Tc) to saturation at 37 °C. About 20 μl of this culture was mixed with 20 μl of a saturated culture of MM294/RK2013[34]. The mixtures were incubated on the surface of an LB plate at 37 °C for 3 h, and then 20 μl of a saturated LB culture of a streptomycin (Sm^r) resistant derivative of M7922 was mixed into the patch of growth. After further incubation of the plate for 3 h, M7922 transconjugants were selected on LB agar plates containing 100 μg ml^{-1} Sm and 15 μg ml^{-1} Tc. These were scored for sensitivity to 30 μg ml^{-1} kanamycin (Km) to assure that these strains did not also inherit pRK2013. Production of the cholera toxin B subunit was performed by growing cells in CYE medium[39] at 37 °C for 18 h with aeration in the presence of 15 μg ml^{-1} tetracycline. B subunit was measured in cell extracts of MS371 derivatives prepared as described previously[19] and in culture supernatants of M7922 derivatives. As in the case of cholera toxin[19], less than 5% of the total B subunit produced was found in the supernatant of MS371 derivatives and less than 1% was found in cell extracts of M7922 derivatives. The B subunit was quantitated as antigen using a GM_1-dependent, enzyme-linked immunosorbent assay[40] and purified toxin as standard. Values are reported as μg ml^{-1} toxin-antigen equivalents, and less than 0.005 μg ml^{-1} could not be detected in the assay. ND, not determined. Due to NIH recombinant DNA guideline restrictions, plasmid pJM17 was not transferred to *V. cholerae* M7922.

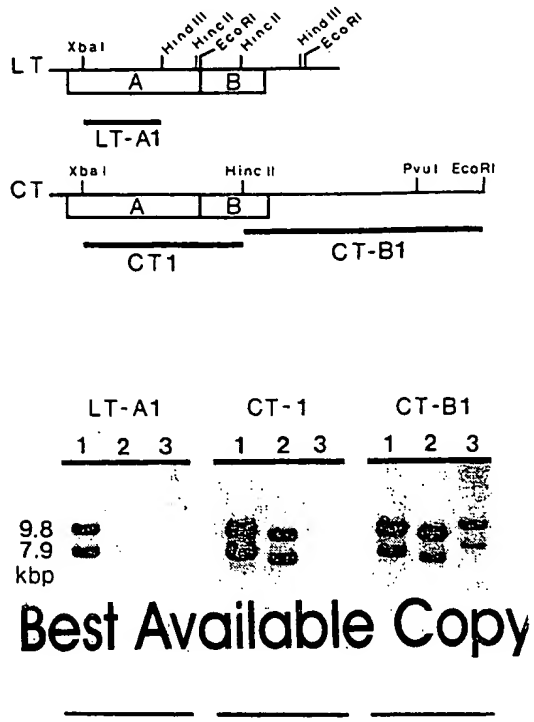

Fig. 5 Southern blot analysis of 0395 and *ctx* deletion derivatives. The top shows partial restriction maps of cloned inserts carrying the *elt* (LT) and *ctx* (CT) operons and the location of restriction enzyme fragments (heavy lines) used as probes. Shown below are exposures obtained after Southern blot analysis of DNA from wild-type and *ctx* mutant strains. The lanes contained DNA from the following *V. cholerae* strains: 1, 0395; 2, 0395-N1; 3, 0395-NT. The more efficient hybridization of the CT-B1 probe to the upper bands in all three lanes reflects the fact that this probe contains not only *ctxB* sequences but also sequences unique to the downstream flanking DNA of the *ctx* locus located on these larger *Xba*I chromosomal fragments. The sizes of the two original wild-type 0395 *Xba*I fragments containing the two *ctx* copies are shown in kilobase pairs (kbp).
Methods: Probe preparation. The LT-A1 probe is a 475 bp *Xba*I–*Hin*dIII fragment from plasmid EWD299[18]; the CT-1 probe is a 918 bp *Xba*I–*Hin*cII fragment and the CT-B1 probe is a 1.45 kbp *Hin*cII–*Eco*RI fragment, both from pJM17[19]. The fragments were purified by electroelution from gel slices and labelled with [α-^{32}P]dCTP (7,000 Ci mmol^{-1}, NEN) by nick translation[36]. DNA from *V. cholerae* strains (1 μg) was digested with *Xba*I, fractionated by electrophoresis in 0.7% agarose gels, and transferred to a nitrocellulose sheet by the method of Southern[52]. The nitrocellulose was then hybridized sequentially with the LT-A1, CT-1 and CT-B1 probes. Between each successive hybridization the displayed autoradiographic exposures (lower panels) were obtained and the previous probe was removed by treatment of the nitrocellulose with 20 mM NaOH for 2 h. Hybridizations were done at 37 °C for 18 h and, depending on the probe being used, were carried out in solutions of either high (CT-1 and CT-B1 probes) or low (LT-A1 probe) stringency. High-stringency solution contained 50% (v/v) formamide, 75 mM Na citrate, pH 7.0, 0.75 M NaCl, 1 mM EDTA, 0.1 SDS, 0.05% bovine serum albumin, 0.05% polyvinyl pyrrolidone, 0.05% Ficoll and 250 μg ml^{-1} heat-denatured calf thymus DNA; low-stringency solution was identical except that it contained 25% (v/v) formamide.

specific deletion mutations in the *ctx* operon. *In vivo* recombination of these mutations into both *ctx* copies of *V. cholerae* 0395 has resulted in the construction of A$^-$B$^+$ (0395-N1) and A$^-$B$^-$ (0395-NT) derivatives of this strain. Given the fact that primary infections with strain 0395 have been shown to induce long-lasting immunity in human volunteers[9,10,17], we anticipate that the derivatives 0395-N1 and 0395-NT can be used as stably attenuated, live oral cholera vaccines. Previous *V. cholerae* toxin mutants developed for this purpose have been less suitable because of factors such as genetic instability of toxin mutations[16] or poor colonization properties[17]. We have presumably avoided these problems by incorporating precise *ctx* deletion mutations into an unmutagenized parental strain with proven human colonization and immunogenic properties. Given that the parental strain 0395 carries two copies of the *ctx* operon, construction of vaccine candidates such as those described here would be essentially impossible by conventional genetic methods involving chemical mutagenesis. Thus, while it is clear that potential hazards are associated with the cloning of toxin genes, the application of this technology to the genetic manipulation of toxins does greatly facilitate the construction of safer and more effective bacterial vaccines.

This work was supported by the Medical Foundation, Inc., Boston, Massachusetts, and grant AI-18045 to J.J.M. from the National Institute of Allergy and Infectious Disease. Work at Smith-Kline-RIT was supported by IRSIA. We thank P. Piot and M. Levine for providing *V. cholerae* strains and D. Portnoy and F. Ausubel for plasmids pRK290, pRK2013 and pHI1. We also thank J. Felton for the critical reading of the manuscript.

Received 6 June; accepted 2 September 1983.

1. Finkelstein, R. A. *CRC crit. Rev. Microbiol.* 2, 553–623 (1973).
2. Collier, R. J. & Mekalanos, J. J. in *Multifunctional Proteins* (eds Bisswanger, H. & Schmincke-Ott, E.) 261–291 (Wiley, New York, 1980).
3. Clements, J. D. & Finkelstein, R. A. *Infect. Immun.* 21, 1036–1039 (1978).
4. Gill, D. M. *Biochemistry* 15, 1242–1248 (1976).
5. Mekalanos, J. J., Collier, R. J. & Romig, W. R. *J. biol. Chem.* 254, 5855–5861 (1979).
6. Gill, D. M. & Merin, R. *Proc. natn. Acad. Sci. U.S.A.* 75, 3050–3054 (1978).
7. Vine, H. L. & Cuatrecasas, P. *Pharmac. Ther.* 12, 167–207 (1981).

8. Levine, M. M. in *Cholera and Related Diarrheas* (eds Ouchterlony, O. & Holmgren, J.) 195–203 (Karger, Basel, 1980).
9. Cash, R. A. *et al. J. infect. Dis.* 130, 325–333 (1974).
10. Levine, M. M. *et al. J. infect. Dis.* 143, 818–820 (1981).
11. Mukerjee, S. *Bull. Wld Hlth Org.* 29, 753–766 (1963).
12. Howard, B. D. *Nature* 230, 97–99 (1971).
13. Finkelstein, R. A., Vasil, M. L. & Holmes, R. K. *J. infect. Dis.* 129, 117–123 (1974).
14. Honda, T. & Finkelstein, R. A. *Proc. natn. Acad. Sci. U.S.A.* 76, 2052–2056 (1979).
15. Mekalanos, J. J., Moseley, S. L., Murphy, J. R. & Falkow, S. *Proc. natn. Acad. Sci. U.S.A.* 79, 151–155 (1982).
16. Woodward, W. E., Gilman, R. H., Hornick, R. B., Libonati, J. P. & Cash, R. A. *Devl biol. Standard.* 33, 108–112 (1976).
17. Levine, M. M. *et al.* in *Acute Enteric Infections in Children. New Prospects for Treatment and Prevention* (eds Holme, T., Holmgren, J., Merson, M. H. & Mollby, R.) 443–459 (Elsevier, Amsterdam, 1981).
18. Dallas, W. S., Gill, D. M. & Falkow, S. *J. Bact.* 139, 850–858 (1979).
19. Pearson, G. D. N. & Mekalanos, J. J. *Proc. natn. Acad. Sci. U.S.A.* 79, 2976–2980 (1982).
20. Dallas, W. S. & Falkow, S. *Nature* 288, 499–501 (1980).
21. Spicer, E. K. & Noble, J. A. *J. biol. Chem.* 257, 5716–5721 (1982).
22. Lai, C. Y., Cancedda, F. & Chang, D. *FEBS Lett.* 100, 85–89 (1979).
23. Mendez, E., Lai, C. Y. & Wodnar-Filipowica, A. *Biochem. biophys. Res. Commun.* 67, 1435–1443 (1975).
24. Kurosky, A., Markel, D. E., Touchstone, B. & Peterson, J. W. *J. infect. Dis.* 133, S14–S22 (1976).
25. Kurosky, A., Markel, D. E., Peterson, J. W. & Fitch, W. M. *Science* 195, 299–301 (1977).
26. Lai, C. Y. *J. biol. Chem.* 252, 7249–7256 (1977).
27. Rosenberg, M. & Court, D. *A. Rev. Genet.* 13, 319–353 (1979).
28. Platt, T. *Cell* 24, 10–23 (1981).
29. Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* 71, 1342–1346 (1974).
30. Sanger, F., Carlson, A. R., Hong, G. F., Hill, D. E. & Peterson, G. B. *J. molec. Biol.* 162, 729–773 (1982).
31. Oppenheim, D. S. & Yanofsky, C. *Genetics* 95, 785–795 (1980).
32. Daniels, D. L. & Blattner, F. R. *Virology* 117, 81–92 (1982).
33. Lai, C. Y., Cancedda, F. & Duffy, L. K. *Abstr. 11th int. Congr. Biochem.*, Toronto, Canada, 207 (National Research Council of Canada, Ottawa, 1979).
34. Ruvkun, G. B. & Ausubel, F. M. *Nature* 289, 85–88 (1981).
35. Guerrant, R. L., Brunton, L. L., Schnaitman, T. C., Rebhun, L. I. & Gilman, A. G. *Infect. Immun.* 10, 320–327 (1974).
36. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning* 413 (Cold Spring Harbor Laboratory, New York, 1982).
37. Baine, W. B., Vasil, M. L. & Holmes, R. K. *Infect. Immun.* 21, 194–200 (1978).
38. Mekalanos, J. J. & Murphy, J. R. *J. Bact.* 141, 570–576 (1980).
39. Mekalanos, J. J., Collier, R. J. & Romig, W. R. *Infect. Immun.* 20, 552–558 (1977).
40. Holmgren, J. *Infect. Immun.* 8, 851–859 (1973).
41. Covarrubias, L. *et al. Gene* 13, 25–35 (1981).
42. Moseley, S. L. & Falkow, S. *J. Bact.* 144, 444–446 (1980).
43. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* 65, 499–560 (1977).
44. Smith, A. S. H. *Meth. Enzym.* 65, 499–560 (1977).
45. Brammar, W. J., Berger, H. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* 58, 1499–1506 (1967).
46. Pribnow, D. *J. molec. Biol.* 99, 419–443 (1975).
47. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* 33, 717–728 (1983).
48. Donahue, T. F., Davis, R. S., Lucchini, G. & Fink, G. R. *Cell* 32, 89–98 (1983).
49. Ruvkun, G. B., Sundaresan, V. & Ausubel, F. M. *Cell* 29, 551–559 (1982).
50. Berlinger, J. E., Beyon, J. L., Buchanan-Wollaston, A. V. & Johnston, A. W. B. *Nature* 276, 633–634 (1978).
51. Grunstein, M. & Hogness, *Proc. natn. Acad. Sci. U.S.A.* 72, 3961–3965 (1975).
52. Southern, E. M. *J. molec. Biol.* 98, 503–517 (1975).
53. Mekalanos, J. J. *Cell* 35 (in the press).

Cell-specific expression controlled by the 5′-flanking region of insulin and chymotrypsin genes

Michael D. Walker, Thomas Edlund, Anne M. Boulet & William J. Rutter

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA

DNA sequences containing the 5′-flanking regions of the insulin and chymotrypsin genes were linked to the coding sequence of the chloramphenicol acetyltransferase (CAT) gene. The insulin gene recombinant elicits preferential expression of CAT activity when introduced into cells producing insulin; similarly, the chymotrypsin gene recombinant elicits preferential expression in chymotrypsin-producing cells. Sequences located upstream of previously defined transcriptional control elements are essential for efficient expression in both cases.

DIFFERENTIATED cells in eukaryotes possess a remarkable capacity for selective expression of specific genes. A single gene may account for a large fraction of the total gene expression in one cell type, yet be expressed at undetectable levels in other cells. The level of control is very large, and may approach a million-fold[1–3]. Current evidence suggests that the rate of transcription is a key aspect in this control of gene expression[4]. In contrast to the great specificity seen *in vivo*, it is possible to detect expression of protein coding genes of higher eukaryotes using a variety of heterologous *in vivo* and *in vitro* systems[5–8]; however, the levels of expression observed are extremely low compared with those found in differentiated cells *in vivo*. Analyses of these activities have revealed the presence of a number of distinct transcriptional control elements located close to the mRNA initiation site. One of these sequences (the TATA or Goldberg–Hogness box) is located about 30 base pairs (bp) upstream from the initiation site. The TATA sequence is present in most, but not all, genes and helps to determine the efficiency and the precise location of the transcription start site[9]. Two less well defined elements are located further upstream: the first occurs 70–80 bp from the transcription start site, and has the consensus sequence CCAAT[10,11], and the second lies 80–110 bp upstream from the transcription start site[11,12]. Careful analyses of the expression of the β-globin gene and the herpes thymidine kinase (TK) gene have failed to reveal a functional role for DNA sequences further upstream than around 110 bases[11,12]. No evidence exists linking these elements to cell-specific expression.

Study of viral transcription units has identified a different class of control element typically located 100–300 bp upstream from the transcription start site. These elements, termed enhancers, are distinguished by their ability to activate transcription units independently of their orientation or precise positioning relative to the unit[13–16]. Enhancers operate in a variety of cell types, although not necessarily with equal efficiency: the polyoma enhancer, for example, is functional in differentiated mouse cells but not in undifferentiated embryonal cells[17,18].

There have been relatively few reports documenting introduction of genes into differentiated cells which express the endogenous cognate gene[19–24]. Stable transformants of myeloma cells express higher levels of a transferred immunoglobulin κ gene than similar transformants of fibroblasts[22]. Expression of the chick δ crystallin gene is higher on microinjection into lens cells than fibroblasts[24].

We have designed experiments to test for elements of genes required for cell-specific expression. These involve linking a putative regulatory control region to an enzymatic 'reporter' function[25]. The genes coding for mammalian insulin[26–29] and rat chymotrypsin B (G. I. Bell, C. Quinto, C. S. Craik and W.J.R., unpublished results) have been cloned and characterized. They are expressed at high level only in the pancreas. However, clearly distinct cell types are involved: insulin is synthesized in endocrine β-cells and chymotrypsin in exocrine cells[30]. We linked DNA sequences containing the 5′-flanking region of the insulin and chymotrypsin genes to the coding sequence of chloramphenicol acetyltransferase (CAT), a particularly useful reporter function, developed by Gorman *et al.*[31]. Insulin and chymotrypsin gene recombinants were introduced into pancreatic endocrine and exocrine cells as well as into non-pancreatic cells, and the relative levels of transient

0028-0836/83/490557—05$01.00